Power and Thermal Challenges in Mobile Devices

Krishna Sekar
Broadcom Corporation
16340 W Bernardo Drive
San Diego, CA, 92127
1(858)385-8846
ksekar@broadcom.com

ABSTRACT

In spite of significant advances in the development of low-power designs and power management techniques, power remains and will remain a first-class design constraint for mobile devices. The functionality integrated into mobile devices will only continue to grow with the increasing transition from PCs to super-phones. Power remains one of the critical hurdles to this transition, and without continued innovations in power management, the term "mobile" devices will be rendered moot. In addition to (and closely related to) power challenges in mobile devices are thermal challenges. Increasingly complex and rich functionality in mobile devices leads to higher power dissipation, and consequently, higher temperatures. However, thermal constraints are constant across successive device generations. Thermal limits, even more than power, will become the fundamental bottleneck to increasing the capabilities of such devices, making thermal management techniques crucial.

This paper discusses some of the major challenges in power management and thermal management for current and next-generation mobile devices from a semiconductor industry perspective. Specifically, this paper discusses challenges in three areas: process-variability-aware power management, thermally aware power management and thermal management for mobile devices. Broadcom mobile chipsets feature multiple such advanced power and thermal management techniques, with active on-going research and development in each of these areas.

Categories and Subject Descriptors

C.5.3 [**Computer System Implementation**]: Microcomputers – *Portable devices*

Keywords

Mobile devices; power management; thermal management; leakage power; process variability

1. INTRODUCTION

High-end mobile computation and communication devices have become ubiquitous today. The research firm IDC reports that for the first time, more smartphones were shipped in Q1 2013 than feature phones [1]. This trend will continue with the increasing availability of sub-$100 smartphones and tablet devices [2]. Significant advances in mobile technology have enabled this transition. Important among them has been the development of advanced low-power designs and power management techniques, allowing increasingly rich and complex feature-sets to be run on devices with only marginally increasing battery capacity. This is illustrated by Figure 1, which shows how the battery capacities of both Apple iPhones and Samsung Galaxy smartphone models have changed over successive generations of these devices [3][4]. Battery capacity has remained roughly flat between the 1st Generation iPhone and iPhone 5, while it has increased only 1.73 times between Samsung Galaxy S and S4. However, performance has increased by many times more. For example, the application processor performance has increased from up to 2000 DMIPS in Galaxy S with a single ARM8 processor at 1GHz [5][6], to up to 31500 DMIPS in Galaxy S4 with quad ARM A15 processors at 1.6GHz and quad ARM A7 processors at 1.2GHz[7][8], an increase of more than 15 times!

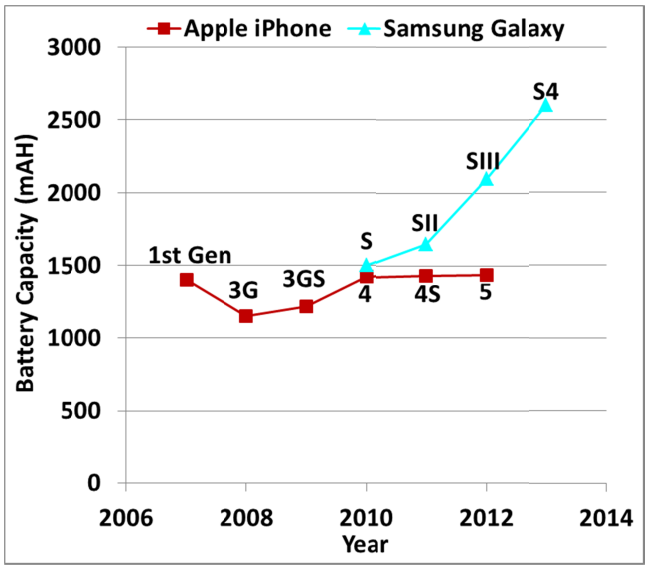

Figure 1: Smartphone battery capacity evolution

In spite of these advances, power remains and will remain a first-class design constraint to the continued advancement of mobile technology. Already users are implicitly expected to charge their mobile devices one or more times a day. Also, there is a limit to the battery size that can be accommodated in ever decreasing device form factors. However, the functionality integrated into such devices, and consequently their power consumption, will continue to grow as users increasingly transition to them from PCs as their primary means of information processing, consumption

MobiCom'13, September 30–October 4, 2013, Miami, FL, USA.

and communication. Without continued innovations in power management, this transition cannot fully come to fruition.

This paper discusses two such power management challenges in mobile devices:

- **Process-variability-aware Power Management:** With shrinking semiconductor process technology nodes, there is much greater variability between the device characteristics of individual dice. A power management policy which is optimized for a "nominal" die may be suboptimal for a "non-nominal" die. However, power management policies today are often process-variability agnostic. This is discussed in Section 2.

- **Thermally Aware Power Management:** Along with device variability, smaller process technology nodes also lead to much higher leakage power in semiconductor devices, making leakage power as important to manage as dynamic power. This is especially true at higher temperatures since leakage power increases exponentially with temperature. The power optimization techniques for leakage power may be very different from those for dynamic power. However, power management policies are usually temperature agnostic. Section 3 further discusses this challenge.

Closely related to, but distinct from, power challenges in mobile devices are thermal challenges. With multiple GHz+ application processors [9], advanced graphics capabilities [10], advanced LTE modems [11] and 5G WiFi [12] technologies, amongst others, modern mobile devices are capable of supporting very sophisticated functionality with high levels of performance. This leads to much higher power dissipation, and consequently, much higher temperatures. However, the thermal constraints of components inside a mobile device remain relatively constant. Furthermore, the thermal constraints on mobile device surface temperatures are also constant since the ability of human skin to tolerate temperature does not change. In fact, the problem is exacerbated with the decrease in the form factors of mobile devices which decreases their ability to dissipate the generated heat. This could lead to device overheating issues with very adverse consequences in terms of device usability, consumer safety, product liability, and negative publicity [13][14]. Thermal constraints, even more than power, will become the fundamental bottleneck to further integrating more functionality and performance in mobile devices.

Thermal challenges are being addressed today through the use of thermal management policies that typically throttle the performance of various components under high temperature conditions. However, without careful design, such policies could force devices to be throttled to a much lower level of performance for significant periods of time, invalidating the whole point of adding such advanced capabilities in the first place. This paper discusses the thermal management challenge in Section 4, advocating the need for optimized thermal management algorithms and policies that can intelligently manage the trade-off between application performance, power dissipation, thermal effects and user-perceived quality.

These challenges are being addressed through advanced power and thermal management techniques in various different Broadcom mobile chipsets today such as cellular baseband processors, mobile multimedia processors, power management ICs, cellular RF ICs, WiFi, Bluetooth, Global Positioning System (GPS), and Near-Field-Communications (NFC) systems, amongst others [15]. Broadcom is also pursuing active research and development in each of these areas.

2. PROCESS-VARIABILITY-AWARE POWER MANAGEMENT

2.1 Description

Continued advances in semiconductor device fabrication technology have enabled progressive reduction of transistor process geometries. Figure 2 shows how the process technology node (referring to the average half-pitch of a DRAM cell) has decreased over the years, from 130 nm in 2001, to 28 nm in 2013, and projected to reach 10 nm by 2022 [16]. This shrinking of transistor process geometries has been the main driving force behind Moore's law [17], which predicts that the number of transistors on integrated circuits doubles approximately every two years. It is the continued applicability of Moore's law that has enabled the revolution in mobile technology with faster, smaller and cheaper devices becoming available every successive generation.

However, with smaller process geometries, there is greater variability in device characteristics such as performance and power between individual dice. For example, in the 28 nm technology node, there could be up to 42% variability in performance and up to 51% variability in total power consumption [18]. This increasing die-to-die variability can be attributed to inherent limitations of the fabrication process especially for smaller process geometries. This trend has only become worse at smaller technology nodes, with increasing number of silicon dice whose characteristics vary from that of "nominal" or "typical" dice. Figure 3 shows the probability density function for the delay through a flip-flop (normalized to 1) fabricated in 65 nm and in 28 nm. At the 28 nm node, there are a larger number of dice whose delay is more than or less than that of nominal dice. Such devices are characterized as either slow or fast (i.e., non-nominal). The same trend applies to power consumption as well.

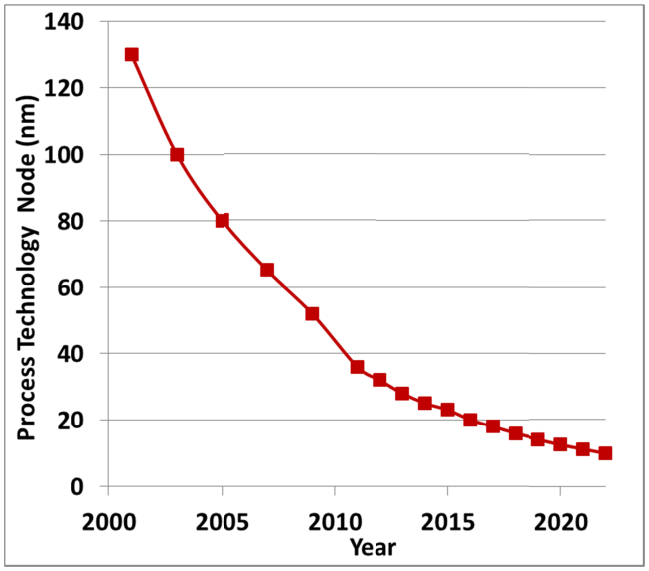

Figure 2: Semiconductor process technology node evolution

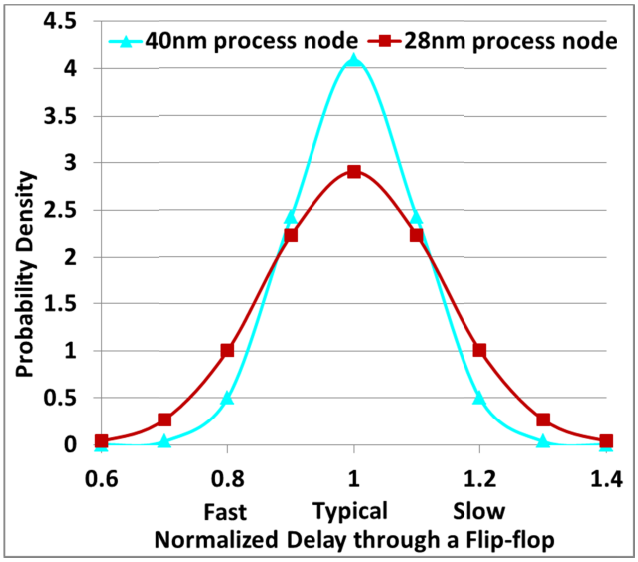

Figure 3: Variability of delay through a flip-flop

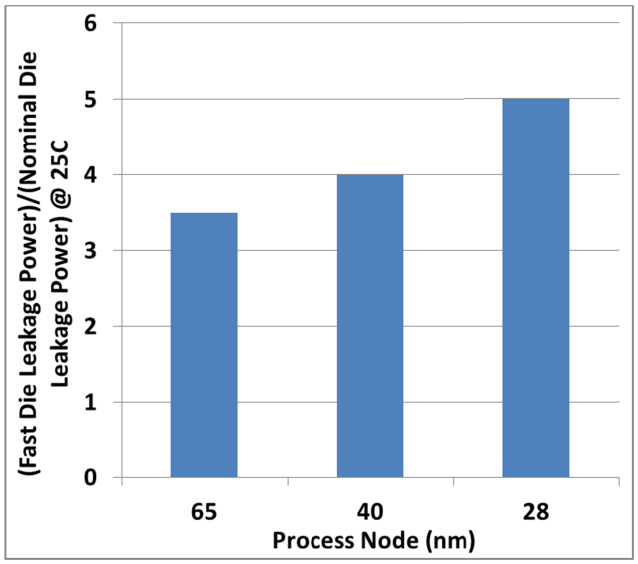

Figure 4: Ratio of the leakage power of fast dice to leakage power of nominal dice for different process technology nodes

To keep silicon yields high and the process node economical, non-nominal dice (within a certain standard deviation such as 2.5σ or 3σ) are still shipped and sold in end products. Therefore, in addition to an increasing number of non-nominal dice, the magnitude of difference between nominal and non-nominal dice has also been increasing with smaller process geometries. Figure 4 shows how the ratio between the leakage power of fast dice and nominal dice has been increasing with smaller technology nodes. The same trend holds true for dynamic power variability as well, albeit to a much smaller magnitude.

The consequence of device power variability is that **power is no longer a number, but a distribution**. However, power models and estimates typically used today are for nominal dice. Also, power management policies used today are usually customized for nominal dice and may not be optimal for the increasingly common non-nominal dice. This calls for *process-variability-aware power management policies*.

2.2 Motivational Example

One such commonly used power management technique is Adaptive Voltage Scaling (AVS). In AVS, slow dice are operated at a higher voltage compared to nominal dice, while fast dice are operated at a lower voltage, such that all dice can achieve a given target clock frequency. This exploits the fact that device speed is a function of the operating voltage. Fast dice, by virtue of being "fast", can still meet timing requirements at a lower voltage, while slow dice can be made to meet timing requirements by increasing their voltage. This, in turn, reduces the power consumption variability between different die types. Fast dice have higher leakage power consumption compared to nominal dice. Hence, reducing their voltage decreases their power consumption, bringing them closer to nominal dice. Similarly, increasing the voltage of slow dice, increases their power consumption, thereby, bringing them closer to nominal dice.

As another example, consider two power management policies for mobile devices: Dynamic Voltage and Frequency Scaling (DVFS) [19] and Run-Fast-and-Stop (RFS). In DVFS, the device is operated at the lowest possible voltage and frequency setting, such that it is just enough to meet its current computational requirements. DVFS operates under the principle that it is more power-efficient to run at the lowest voltage and frequency since, although it increases device active time, it reduces the overall power consumption due to the square dependence of dynamic power on voltage. DVFS is quite commonly used today in mobile devices. In contrast, in RFS, the device is operated at the highest possible voltage and frequency setting, such that the tasks are completed as soon as possible, and then the device is powered off. RFS operates under the principle that it is more power-efficient to keep the device powered off for as long as possible in order to minimize leakage power. These are two very different policies, and yet, each may be optimal for different die types. DVFS may be better suited for nominal and slow dice where dynamic power tends to dominate overall power consumption, whereas RFS may be better suited for fast dice where leakage power may dominate.

2.3 Challenge

A one-size-fits-all power management policy does not work very well any more, necessitating the need for process-variability-aware power management policies that are customized to the device characteristics they run on. Broadcom mobile chipsets feature such advanced power management policies, and Broadcom is actively engaged in further research and development in this area.

3. THERMALLY AWARE POWER MANAGEMENT

3.1 Description

As transistor process geometries have continued to shrink (Figure 2), the leakage power dissipated by semiconductor devices has increased exponentially. This is illustrated by Figure 5, which shows that the leakage power per gate increased by 2x going from 65 nm to 40 nm, and by 3x going from 40 nm to 28 nm. This trend is driven by multiple factors including thinner gate oxides, smaller channel lengths and smaller threshold voltages, and is expected to continue as process geometries shrink further. Additionally, the leakage power of fast dice increases much faster compared to that of nominal dice at smaller process technology nodes (Figure 4). In contrast, dynamic power per gate (for the same frequency) has decreased in every successive technology generation, due to reduction in the device operating voltage and load capacitances.

Therefore, leakage power is becoming a much more dominant factor in the total power dissipation of semiconductor devices.

This is especially true at higher temperatures since leakage power depends exponentially on temperature. Figure 6 shows how the leakage power of a nominal device in 28 nm scales with temperature, from 1x at 25°C to 30x at 125°C. At higher temperatures, leakage power could account for 40%-50% of the total power consumption of a device. Also, at smaller process geometries, dynamic power consumption also increases at higher temperatures although to a much lesser extent compared to leakage power. Furthermore, due to the increasingly rich and power-hungry computation and communication functionality integrated in small form-factor mobile devices today, device temperatures are expected to be significantly high for heavy-usage scenarios.

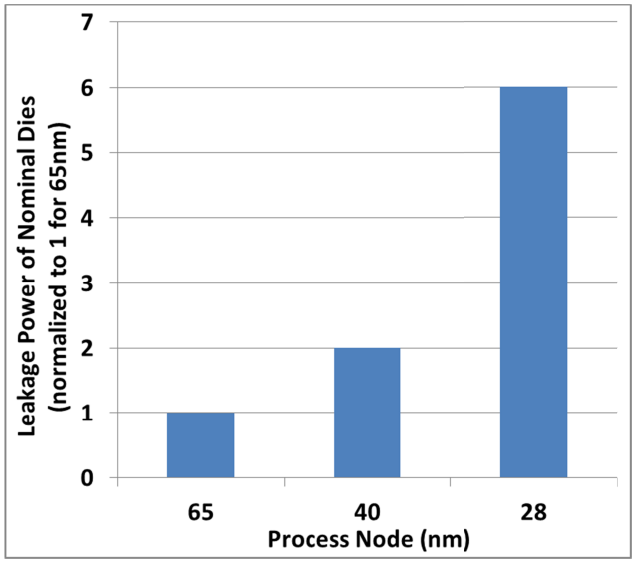

Figure 5: Leakage power of nominal dice (normalized to 1 for 65 nm) for different process technology nodes

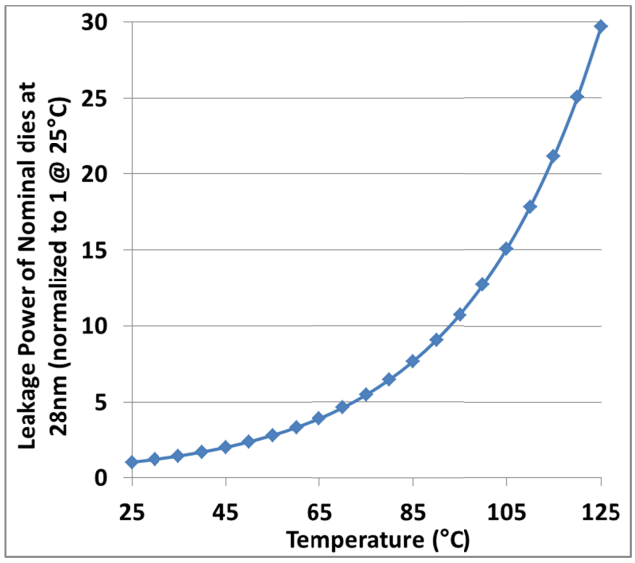

Figure 6: Leakage power of nominal dice (normalized to 1 at 25°C) over temperature

The consequence of the above is that **temperature has become a very important factor to consider in the context of power**. However, power models, simulations and estimates used today usually do not factor in temperature as a parameter. In fact, since temperature is itself a function of the power dissipation of a device, there is a positive feedback loop between power and temperature until the device reaches steady-state temperature (or in worst-case scenarios, there is thermal runaway). This calls for a "power-temperature cosimulation" methodology for accurate power estimation. Furthermore, power management policies typically used in mobile devices today are temperature agnostic.

3.2 Motivational Example

Consider, for example, two CPU power management policies used today in multi-CPU mobile devices: Dynamic Voltage and Frequency Scaling (DVFS) [19] and CPU Hotplug [20]. As described in Section 2.2, in DVFS, the CPU is operated at the lowest voltage and frequency setting which is just enough to meet the CPU load requirements. In contrast, in CPU Hotplug, one or more of the CPU cores are powered down or powered up depending on the CPU load requirements. DVFS is better optimized for dynamic power, while CPU Hotplug is better optimized for leakage power. Now, consider a dual-core mobile device. At low temperatures, due to the dominance of dynamic power, DVFS would be more power-efficient than CPU Hotplug; for processing the same CPU load, it may be better to operate two cores at a lower voltage and frequency than one core at a higher voltage and frequency. However, at high temperatures, due to the dominance of leakage power, the situation may completely reverse and CPU Hotplug may be more power-efficient.

3.3 Challenge

Temperature must be factored in as a critical parameter for all aspects of power on mobile devices, from power modeling, to estimation, to simulation and to thermally aware power management policies. This is the strategy adopted in Broadcom mobile chipsets, where temperature is incorporated as a first-class design parameter into all aspects of mobile device power modeling and management.

4. THERMAL MANAGEMENT

4.1 Description

Over the last several years, there has been an explosive growth in mobile technology, with ultra-high-end mobile devices becoming ubiquitous today. Such devices integrate a wide variety of sophisticated and high-performance computation and communication technologies, such as multiple GHz+ application processors, advanced LTE modems, advanced multimedia and graphics cores, 5G WiFi, GPS, Bluetooth and NFC chipsets, amongst others [15]. Now, for every successive technology node generation, the power dissipation per silicon gate has been decreasing. Therefore, the power consumption for the same use-case has been decreasing. However, due to the integration of more and more rich functionality (with a consequent increase in the number of gates), and increases in the operating frequencies to support such functionality, the overall power dissipated inside mobile devices has been increasing rapidly. Figure 7 illustrates how the power dissipation (normalized to 1W for Chip1) of some representative Broadcom mobile baseband processors has increased over the years for the heaviest usage scenario that they can respectively support and without any thermal management. From a 65 nm chip featuring a single ARM11 application processor, to a 28 nm chip featuring multiple advanced ARM application processors, the potential power dissipation has

increased by almost 3 times. This trend is only expected to continue with further integration of even more advanced functionality in mobile devices.

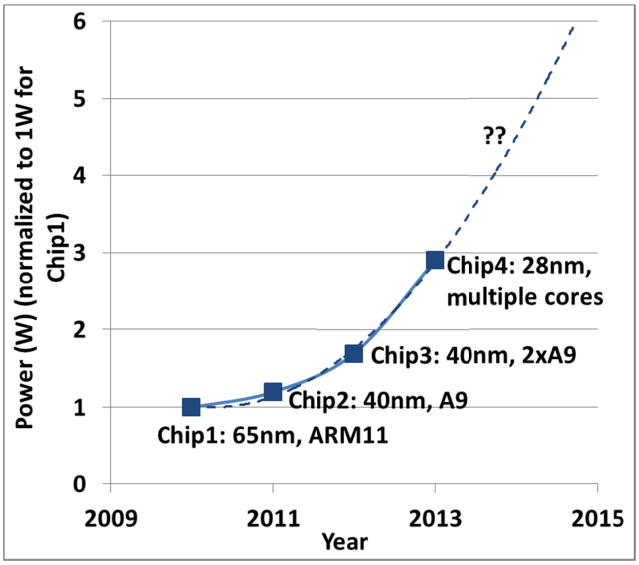

Figure 7: Power dissipation of different representative cellular baseband processors under heaviest usage scenario (normalized to 1W for Chip1)

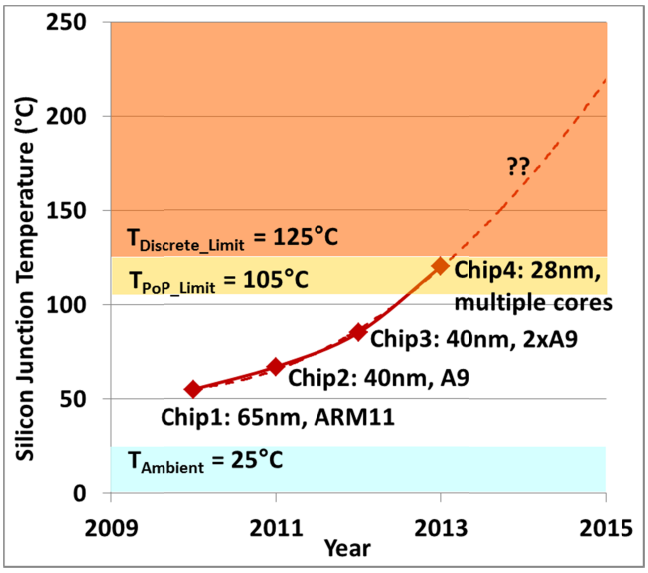

Figure 8: Silicon junction temperature of different representative cellular baseband processors under heaviest usage scenario

Due to increasing power dissipation, the temperatures inside mobile devices have been rapidly increasing. Figure 8 illustrates how the silicon die temperature has increased over the years corresponding to the same chips and power dissipations shown in Figure 7 (JEDEC thermal analysis assuming an ambient temperature of 25°C; without any thermal management). However, the fundamental thermal limits of components inside a mobile device have not changed. For example, in the elevated temperature range, low-power DRAM (LPDDR) devices have a thermal limit of 105°C [21], thereby limiting the maximum temperature of the underlying baseband processors in the case of Package-on-Package (PoP) systems. Similarly, the maximum baseband processor die junction temperatures are usually limited to 125°C to ensure correct device operation. Additionally, the thermal limits on mobile device surface temperatures have remained constant over the years since the ability of human skin to withstand temperature does not change. Furthermore, active cooling mechanisms such as fans, which are commonly used in laptops and PCs for thermal mitigation, are not feasible for use in phones and tablets. The thermal problem is further exacerbated by the fact that mobile device form-factors are becoming smaller and thinner, and therefore their ability to dissipate the generated heat has been decreasing.

One way to mitigate the thermal challenges in mobile devices is through better platform-level thermal design. This includes techniques such as thermally aware component placement and the use of materials such as heat spreaders and thermal interface material (TIM). However, such techniques provide limited improvements and are relatively expensive to use. Temperature, even more than power, is fast becoming the fundamental design bottleneck to the continued progress towards even more powerful mobile devices. Therefore, the design and development of better **thermal management** algorithms and policies has become imperative.

Thermal management policies typically rely on throttling the performance or functionality of various components in order to reduce their power dissipation, and consequently the temperature of mobile devices. However, by constraining the capabilities of the mobile device, thermal management will lead to some loss of performance. Therefore, without careful design of such thermal management policies, mobile devices could end up operating in much lower modes of performance than they are actually capable of for significant periods of time, thereby negating the usefulness of adding such high-performance features in the first place.

4.2 Motivational Example

Consider, for example, two thermal management policies and their effect on power and temperature as illustrated in Figure 9. Assume that the maximum temperature limit of the device is 100°C and the ambient temperature is 25°C. Under Policy1, the device is allowed to dissipate 3W of power until the temperature reaches the maximum limit of 100°C. At this point, the device is throttled to dissipate only 1W of power until the temperature cools down to 90°C, at which point throttling is disengaged. Figure 9 shows the power dissipation (Primary Y-axis, solid red line) and temperature (Secondary Y-axis, dotted red line) of the device over time under Policy1. The long-term average power dissipation of the device under Policy1 is 2.3W. In contrast, under Policy2, once the device reaches a temperature greater than 90°C, it is constantly throttled to dissipate only 2.3W of power. Figure 9 shows the power dissipation (solid blue line) and temperature (dotted blue line) of the device under Policy2. Both Policy1 and Policy2 dissipate the same amount of average power (2.3W), but under Policy2, the device temperature reaches only 95°C, and therefore has 5°C more thermal headroom. This thermal headroom can be translated to allowing the device to dissipate higher power than Policy1 (2.5W, 8.7% higher) while still keeping the temperature below the thermal limit of 100°C. Additionally, the effect of Policy1 and Policy2 on user-perceived application behavior could be very different. Policy1 could lead to visibly jerky application behavior due to the constant toggling between high and low performance modes, whereas Policy2 could lead to smoother, albeit lower, application performance.

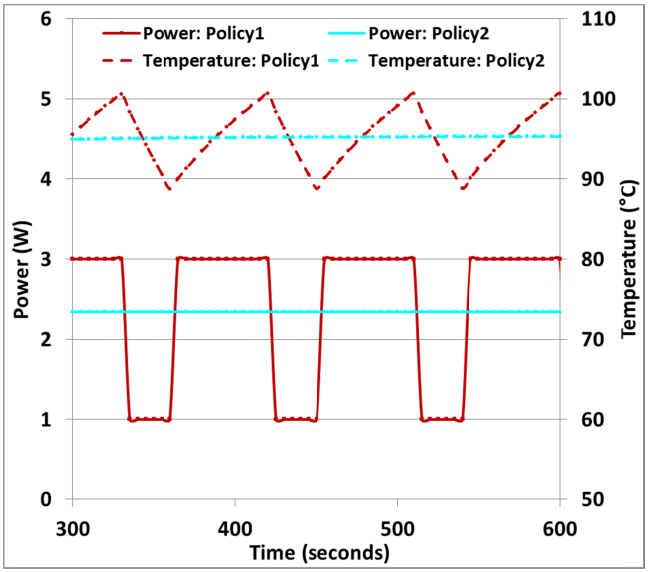

Figure 9: Power and temperature under two different thermal management policies

4.3 Challenge

It is of paramount importance to address the thermal challenge in mobile devices through the design of smart thermal management algorithms and policies that can effectively manage the trade-off between performance, power, temperature, and user experience. Such advanced thermal management features are available on Broadcom mobile chipsets today, with further research and development in this area being actively pursued.

5. CONCLUSIONS

Significant challenges remain today in the areas of power and thermal management for the continued advancement towards ever more sophisticated and feature-rich mobile computation and communication devices. This paper discussed three such challenges in the area of power management and thermal management. Broadcom is at the forefront of addressing many of these power and thermal challenges in various leading-edge mobile chipsets through innovate technologies, thereby enabling the continuation of the mobile revolution.

6. REFERENCES

[1] "More Smartphones Were Shipped in Q1 2013 Than Feature Phones, An Industry First According to IDC", IDC – Press Release, April 25th, 2013, http://www.idc.com/getdoc.jsp?containerId=prUS24085413

[2] "The global smartphone tipping point: '$100 or less'", CNET news article, February 29th, 2012, http://reviews.cnet.com/8301-13970_7-57387397-78/the-global-smartphone-tipping-point-$100-or-less/

[3] "List of iOS devices: iPhone", Wikipedia, http://en.wikipedia.org/wiki/List_of_iOS_devices#iPhone

[4] "Samsung Galaxy: Smartphones", Wikipedia, http://en.wikipedia.org/wiki/Samsung_Galaxy

[5] "Samsung Galaxy S", Wikipedia, http://en.wikipedia.org/wiki/Samsung_Galaxy_S

[6] "Cortex-A8 Processor: Specifications: Dhrystone Performance" http://www.arm.com/products/processors/cortex-a/cortex-a8.php

[7] "Samsung Galaxy S4", Wikipedia, http://en.wikipedia.org/wiki/Samsung_Galaxy_S4

[8] "List of ARM microprocessor cores: Cortex-A7 and Cortex-A15", Wikipedia, https://en.wikipedia.org/wiki/List_of_ARM_microprocessor_cores

[9] 'Broadcom 3G Baseband Processors", http://www.broadcom.com/products/Cellular/3G-Baseband-Processors

[10] "Broadcom Mobile Multimedia Processors", http://www.broadcom.com/products/Cellular/Mobile-Multimedia-Processors

[11] "Broadcom 4G Baseband Processors", http://www.broadcom.com/products/Cellular/4G-Baseband-Processors

[12] "Broadcom 5G WiFi", http://www.broadcom.com/products/features/5gwifi.php

[13] "Our test finds new iPad hits 116 degrees while running games", Consumer Reports, March 20th, 2012, http://news.consumerreports.org/electronics/2012/03/our-test-finds-new-ipad-hits-116-degrees-while-running-games.html

[14] "Does the new iPad have an overheating problem?", CNN, March 21st, 2012, http://www.cnn.com/2012/03/20/tech/gaming-gadgets/ipad-overheating

[15] "Broadcom Products", http://www.broadcom.com/products/

[16] "ITRS 2011 Edition: Lithography: 2011 tables: Table LITH2", International Technology Roadmap for Semiconductors, http://www.itrs.net/Links/2011ITRS/Home2011.htm

[17] Moore, Gordon E. (1965), "Cramming more components onto integrated circuits", in *Electronics*, Volume 38, Number 8, April 19th, 1965.

[18] "ITRS 2011 Edition: Design: 2011 tables: Table DESN10", International Technology Roadmap for Semiconductors, http://www.itrs.net/Links/2011ITRS/Home2011.htm

[19] A. R. Chandrakasan and R. W. Brodersen, *Low Power Digital CMOS Design*, Kluwer Academic Publishers, Norwell, MA, 1995.

[20] "CPU Hotplug", Linaro, https://wiki.linaro.org/WorkingGroups/PowerManagement/Doc/Hotplug

[21] "Low Power Double Data Rate 3 SDRAM (LPDDR3)", JEDEC LPDDR3 Standard, JESD209-3, May, 2012, http://www.jedec.org/standards-documents/docs/jesd209-3